DUPLICATE

Forecasting Development

Forecasting Research Technical Report No. 376

A COMPARISON OF CYCLONE SPECTRA IN FORECASTS FROM THE OPERATIONAL (G1) AND NEW DYNAMICS' TRIAL (NT) VERSIONS OF THE UNIFIED MODEL – AUG to DEC 2001

March 2002

T. Hewson

© Crown Copyright 2002

Met Office , Forecasting Products , Room R203 , London Road , Bracknell , Berkshire ,RG12 2SZ, United Kingdom

00/0486

Forecasting Research
Technical Report No. 376

A COMPARISON OF CYCLONE SPECTRA IN FORECASTS FROM THE OPERATIONAL (G1) AND NEW DYNAMICS' TRIAL (NT) VERSIONS OF THE UNIFIED MODEL – AUG to DEC 2001

by

T. Hewson

March 2002

The Met Office
Forecasting Development
Room R307
London Road
Bracknell
RG12 2SZ
United Kingdom

© Crown Copyright 2002

Permission to quote from this paper should be obtained from the above Met Office division

Please notify us if you change your address or no longer wish to receive these publications.

Tel: 44 (0)1344 856245 Fax: 44 (0)1344 854026 email: jsarmstrong@meto.gov.uk

A COMPARISON OF CYCLONE SPECTRA IN FORECASTS FROM THE OPERATIONAL (G1) AND NEW DYNAMICS' TRIAL (NT) VERSIONS OF THE UNIFIED MODEL – AUG to DEC 2001

Tim Hewson. 8th February 2002. Revised 8th March 2002.

Abstract

This report uses new techniques to assess and compare systematic errors in two fundamentally different versions of the Unified Model. We focus on the practically important handling of surface cyclones, looking for drifts and step changes, with forecast lead time, in the spectrum of cyclones represented by the model integrations over an extended North Atlantic area. Special emphasis is placed on the more extreme cases. In total about 300 separate real-time forecasts have been used, covering the period August to December 2001. The analysis relies on a 'cyclone database': this is essentially a simple spreadsheet in which each row represents a surface cyclonic feature in the model, and each of over 100 columns represents a cyclone attribute, such as position, surface pressure or maximum wind within a 600km radius.

The current operational model shows only small drifts in its cyclone spectra with forecast lead time, the most noticeable one being a tendency to under-represent the number of weaker cyclones, the shortfall being about 15% by T+144. A similar deficiency was also seen in 2000. The new dynamics model version also exhibits the 15% shortfall. In addition it tends to underpredict the number of more intense cyclones, if one's measure of intensity is based on local maxima in the near-surface wind speed field. A marked reduction is seen between T+0 and T+6, suggesting, perhaps, a spin-up problem. These reductions may be due to forecast cyclones being insufficiently asymmetric. Aspects of the New Dynamics model which may be contributing to the above include the level of diffusion, and some intrinsic smoothing within the semi-Lagrangian scheme.

The slightly less favourable results found here for the New Dynamics runs are countered by an improvement being seen, relative to the operational model, in other more traditional rms-type statistics (reported elsewhere).

A COMPARISON OF CYCLONE SPECTRA IN FORECASTS FROM THE OPERATIONAL (G1) AND NEW DYNAMICS' TRIAL (NT) VERSIONS OF THE UNIFIED MODEL – AUG to DEC 2001

Tim Hewson. 8th February 2002. Revised 8th March 2002.

Over recent years a suite of software has been developed for automatically pinpointing a wide range of cyclonic features in numerical model analyses and forecasts. The software records each such feature, along with over 100 diagnostic attributes, as an item in a 'cyclone database'. This study uses the software to compare the representation of cyclones in two Unified Model versions – the New Dynamics Trial version (UM 5, or 'NT') and the operational version (UM 4, or 'G1') - over a contiguous period beginning in August 2001. Whilst standard 'NWP indices' constructed for this period suggested that NT was as good as or better than G1, there was always the likelihood that handling of certain types of cyclone differed from this signal. Of particular interest was the extreme, high profile end of the cyclone spectrum, because of the associated hazardous/severe weather. Thus the requirement arose to adopt a cyclone-centred means of verification, which is the focus of this study. The results can be seen as an important complement to the 'NWP index'. In short, the methodology represents a new and innovative means for evaluating numerical forecasts which links closely to the aims of operational forecasting.
In section 1 the results of an earlier year 2000 analysis for G1 are summarised, whilst section 2 describes the 3-d model datasets available for the Aug-Dec 2001 NT versus G1 trial period. Section 3 then examines the two new cyclone databases constructed for the trial period, highlighting any tendencies seen, with forecast lead time, in cyclone intensity and frequency, for certain cyclone classes, with emphasis on the more extreme cases.

1. Introduction - cyclone representation in operational runs (G1) in 2000

The cyclonic features identified range from relatively weak perturbations known as *potential* (frontal) *waves*, through to more substantial *frontal wave cyclones*, and finally to '*barotropic lows*' - namely low-pressure centres around which there is no significant low level thermal gradient. Analysis here and in section 3 focuses on a database subset which includes only the significant frontal wave cyclones and barotropic lows (collectively referred to now as 'cyclones'). The identification methodology builds initially on objective front techniques described in Hewson (1998a). The cyclone algorithms are covered more thoroughly in Hewson (1997), (1998b) and (2001a). The latter report provides a tabulated summary of the equation sets, an up-to-date summary of diagnostics stored and early results of an analysis of cyclone database data for the very cyclonic year 2000, which is summarised below. The formulation of G1 in 2000 was broadly the same as in the trial period in 2001, although model upgrades took place during both years.

Key Results:

Using, as a measure of intensity, the maximum wind speed in the model, 1km above the earth surface, within a set distance of the cyclone centre (300km or 600km), we found, for G1:

1. No clear evidence of a systematic under- or over-intensification of the more intense cyclones.
2. Clear evidence of a systematic under-prediction of the number of weaker cyclones. A shortfall of about 25% developed between T+0 and T+48.
3. Some evidence that the shortfall for weaker cyclones was focussed on certain geographical regions. Similarly there was some evidence that in (fewer) other regions weaker cyclones were over-predicted.

Figure 1, relating to item 3 (and not included in Hewson(2001a)) suggests that the prediction shortfalls occurred mostly near to south-facing coasts (eg Western Mediterranean), around large lakes (eg Great Lakes in the US), and close to areas characterised by large sea surface temperature (SST) gradients (eg south of Newfoundland). Each such region will at times exhibit large, very localised surface temperature gradients which will be greatly smoothed by the model resolution, but which perhaps contribute through upscale effects to modest cyclogenesis. Note

that the majority of analysed lows in these regions are classed as frontal waves, not barotropic lows (see Figure in Hewson (2001a)) which would be consistent with the deficiency being more in incipient frontal waves rather than in mature barotropic systems. Away from the aforementioned regions Figure 2 also suggests there may also be a shortfall due to insufficient development and/or insufficient deceleration towards the end of the typical north Atlantic storm track (Iceland to Scotland to southern Scandinavia), which in turn may also relate to the NW-SE oriented band of over-prediction seen even further northeast.

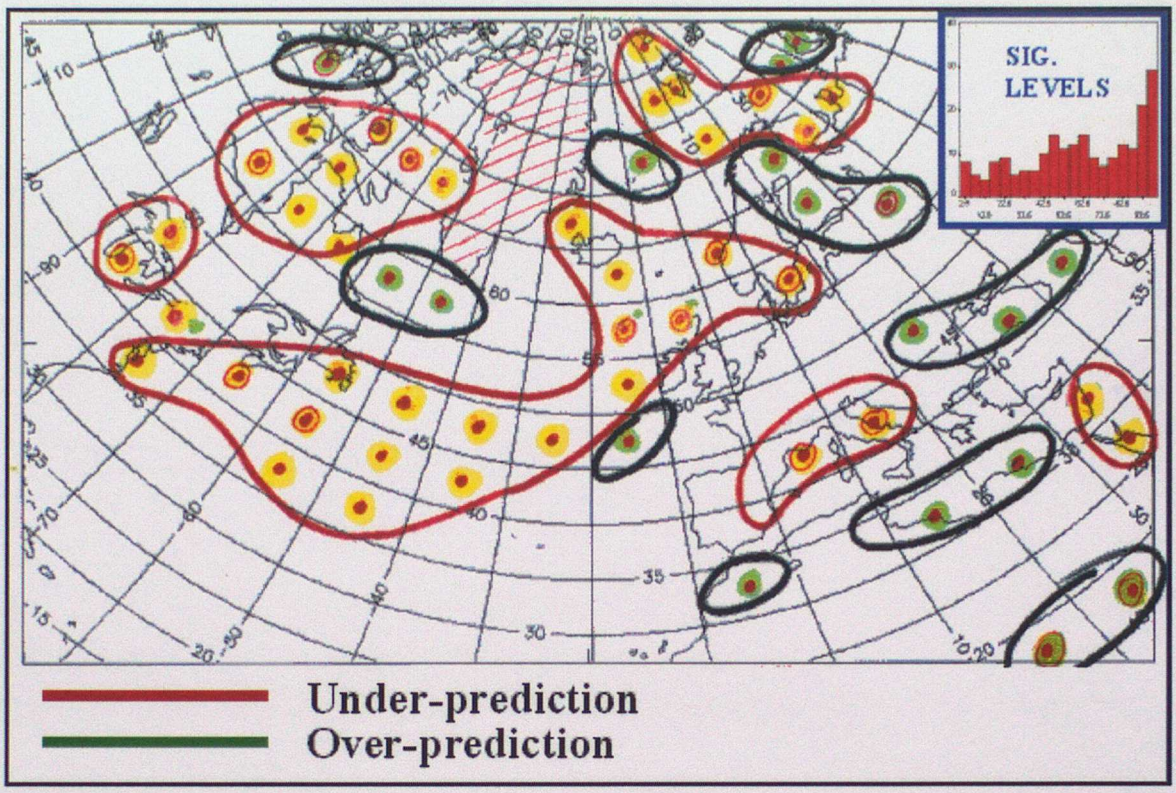

Figure 1: Trends, in aggregated 5° latitude x 10° longitude boxes, in number of cyclones between T+0 and T+144 in G1 during 2000. Yellow spots and red curves denote a noticeable downward trend in numbers – ie under-prediction, whilst green spots and green curves denote an upward trend – ie over-prediction. Circled spots show where the trend was most striking, and thus probably most significant. A bulk test of significance was performed - the inset is a histogram of significance levels (0 to 100%) of differences from a zero trend, obtained across all boxes, when applying simple linear regression to number versus lead time. The increase with x suggests high significance levels are not all occurring just by chance.

2. Data coverage during the trial period August to December 2001

The two input model datasets covering the trial period came from different sources, and had slightly different characteristics:

• The operational (= UM4 = G1) runs were of the same type as used for the 2000 analysis (referenced in section 1). Input data arrived at full horizontal resolution, on all model levels, before being interpolated to a 50km rotated lat-long grid, at 100mb interval pressure levels, covering the old LAM (Limited Area Model) domain (seen in Figure 1). Temporal spacing was regular, being 12-hourly from T+0 through to T+144. One run per day was available, for an 00Z data time.

• The New Dynamics Trial (= UM5 = NT) run data was extracted from the on-line RAID disk, and a dat-tape archive of this data. Each stores input data at full horizontal resolution, but only on standard pressure levels (1000, 850, 700, 500, 300, 200 and 100mb). Interpolation converted this data into an identical (LAM) format to that described above for G1. Temporal spacing was irregular, time frames being T+0, T+6, T+12, T+24, T+48, T+72, T+96, T+120 and T+144. Only one run per day was available for the early part of the period (data time 12Z), but from October 10th 00Z run data was also used. The New Dynamics code underwent several

changes during the analysis period: using NWP nomenclature the model versions used were G4.3 (1/8 – 21/8), G4.4 (22/8-12/9), G4.5 (13/9-7/11) and G4.6 (8/11-18/12), whilst the two main versions of the assimilation scheme used were 18/1.3 (1/8-1/10) and 18/1.4 (2/10-18/12). A description of the changes is beyond the scope of this study, but can be referenced locally at http:///www-nwp/~frgg/EXPTS/um5p2-expts4p9.html and links. Diffusion (∇^4 in winds, theta and moisture, with an e-folding time of 4 hours) remained constant during the period.

The key difference above is that model levels were used for G1 and standard pressure levels for NT. The lower vertical resolution for NT data will impact on certain diagnostics in the database, especially the 'wind speed 1km above the earth's surface', which forms the basis for Figures 4 and 6, which will tend to exhibit lower values for NT than for G1 because of the logarithmic, not linear, wind speed profiles that prevail in the boundary layer. This means that a *direct* comparison of G1 and NT data is not possible, and should not be attempted on any figures that follow in section 3. However it is still valid to compare trends between the models' forecasts and their own respective analyses. This is the approach adopted here.

Table 1 shows input data availability by date. Missing runs – white squares – are mostly attributable to power outages or workstation failures. T+0 and T+6 frames in the NT data were not archived for the period 19th December to 13th January 2002; for this reason the analysis period ends on 18th December. Whilst there is not a direct one-to-one correspondance between the two datasets, they do cover essentially the same period, albeit with a slant towards winter conditions in the NT runs due to introduction of the 00Z run mid-way. No corrections were made to remove forecast frames for which there was no corresponding analysis, nor vice versa, because missing data occurred randomly, and because the dataset was considered sufficiently long for any missing data to have, overall, a negligible impact.

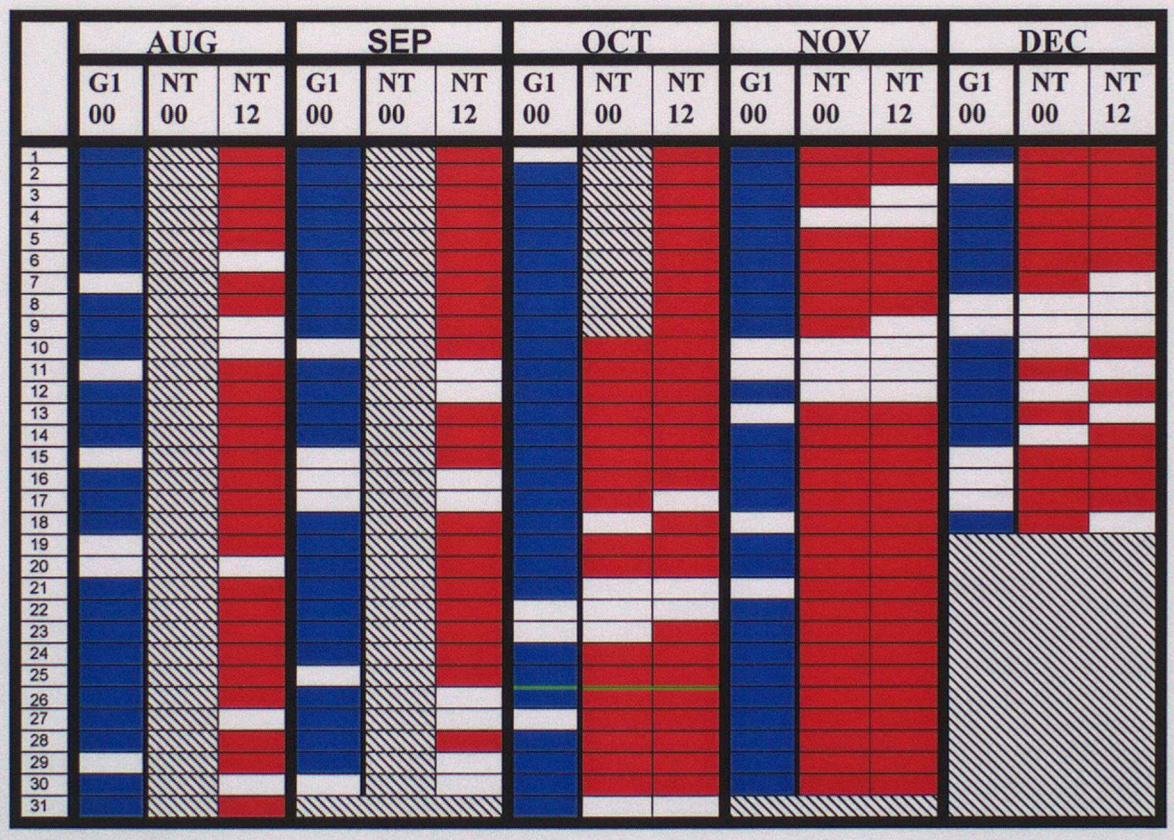

Table 1: data coverage by date for Operational (= G1 = UM4, blue) and New Dynamics trial (= NT = UM5, red) runs during 2001. White squares denote missing runs. Numbers in the second row show analysis clock time (UTC).

Two other point arises in connection with the NT data:

a) There appear to be some weaknesses in the way that 1000mb (and occasionally 850mb) level data is derived when the level is below ground. Downward extrapolation is necessarily used; for geopotential height this involves selecting the first model level that is more than 2km above the lowest model level. A hypothetical temperature structure below this is then derived, using a standard lapse rate down to the lowest model level, and the altimeter equation from

here to the pressure in question. Geopotential is then derived hydrostatically. In the author's view this is a little haphazard - 2km seems rather high, whilst the standard lapse rate assumption must be inappropriate at times. Indeed it was not uncommon for the objective techniques, which use 1000mb geopotential as a proxy for surface pressure, to mis-identify isolated mountain peaks as 'barotropic lows' at times when the background gradient was relatively weak. This is not generally seen in the G1 data. Fortunately the problem has been mitigated somewhat by the fact that the NT orography is much smoother than the G1 orography; in particular isolated peaks are significantly lower.

b) Covariance statistics used in assimilation in the NT runs were not the same as in the G1 runs during the trial – they had previously been re-computed. This, together with the fact that visual inspection of overlayed G1 and NT mslp T+0 fields, on a near-daily basis, showed very few occasions of disagreement, suggests that the G1 analyses should be reasonably reliable.

Late in 2001 and early in 2002 the cyclone database software was applied to the input data described above, creating two new cyclone databases. Each contained approximately 75,000 instances of cyclonic features - which is about 1 for every 6 of the lat-long boxes shown on Figure 1 on each time frame - although then removing weak frontal waves and potential waves reduced case numbers by about 65%. Running as sole job on an HP C360 workstation the database processing had progressed at a rate of about 15 to 25 forecasts per day. A fli file animation for each forecast has been saved on-line, as a reference point. Each animation shows mslp, objective fronts, and each cyclonic feature stored in the database – Figure 5 is an example snapshot.

3. Cyclone spectra during the trial period

This section compares and contrasts the cyclone spectra seen in the G1 and NT runs. Most plots show 'lead time' as the abscissa in order to focus on trends during the forecast period; for G1 this scale is linear, for NT it is not. Note also that the larger number of cases frequently seen on NT plots is due to the larger number of NT forecasts (Table 1).

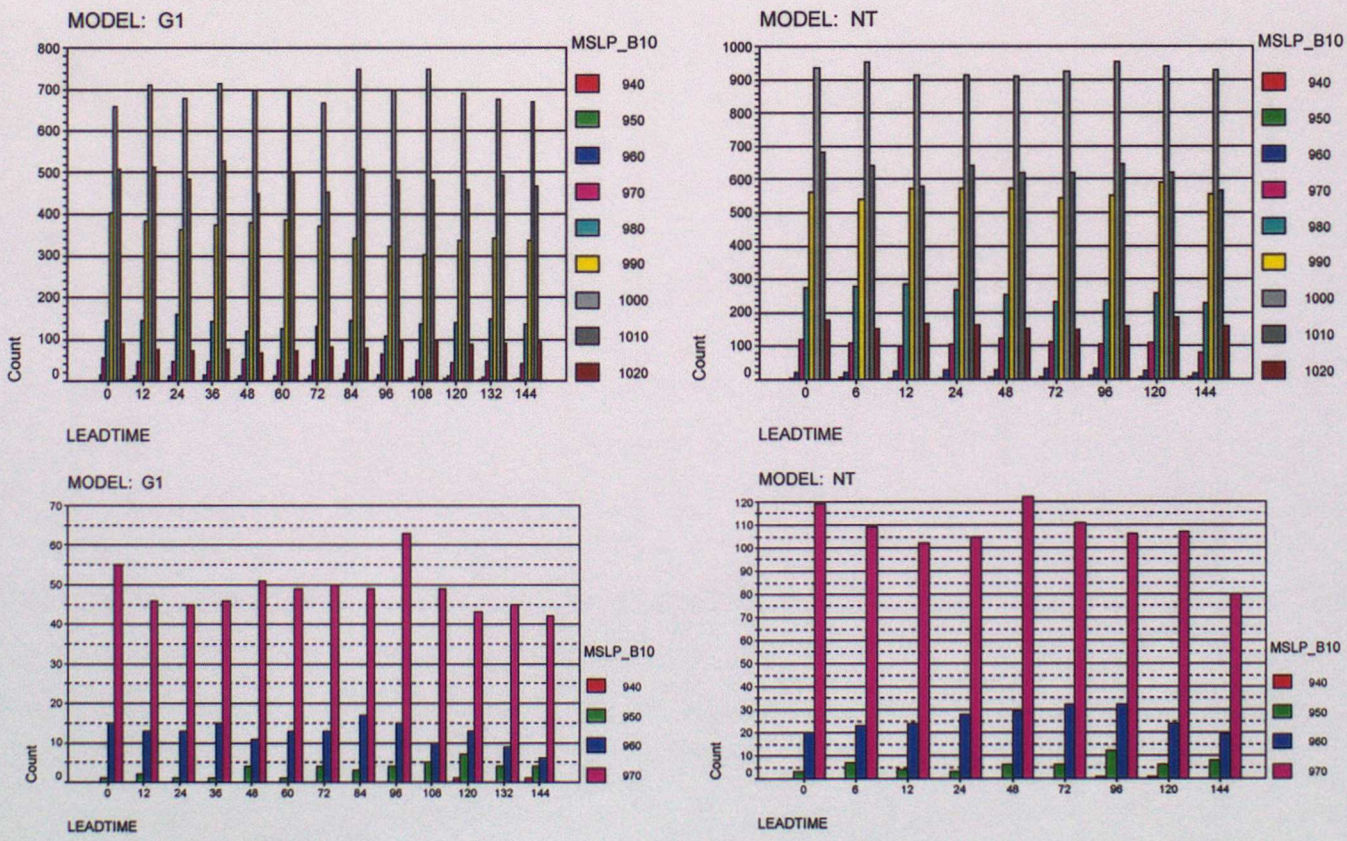

Figure 2: Spectra of cyclone central pressures in G1 and NT runs, at different lead times, over the full LAM domain. Top panels show all cyclones. Lower panels are expanded equivalents of the above, showing only the deepest cyclones, with pressure below 970mb. Legend numbers are the lower limits of 10mb ranges.

Figure 2 shows the depths of all cyclones. For both models there is an erratic but slow reduction with lead time in the total number of what may be termed 'moderate' and 'deep' cyclones, say with central pressures of 970 to 1020mb. For G1 the reduction is most clearly seen in the pink, yellow and dark grey categories, whilst for NT it is in the pink, light blue and dark grey categories, although this category difference is probably not significant. For very deep cyclones, say below 960mb (red and green) the main signal seems to be a tendency to over-deepen (or possibly over-represent) - ie there are fewer very deep cyclones in both models' analyses than in their forecasts – 1 analysed in G1 compares to an average of about 4 in day 4 to 6 forecasts, and 3 analysed in NT compares with about 8 in day 4 to 6 forecasts. However as usual when dealing with extreme cyclones, insufficient sampling is a potential problem here. The data gaps (white boxes on Table 1) may also have had an adverse impact.

Whilst it is common and relatively straightforward to analyse cyclone central pressures, this is not a key parameter as far as weather forecasting is concerned; rather there tends to be an implicit assumption that deeper means more intense, and therefore stronger winds and more rain. Of course in selecting an analysis method there has to be a trade off between simplicity and relevance. Conveniently, the cyclone database not only allows other more relevant parameters to be investigated, but also permits relationships between different parameters to be assessed. Figure 3 shows, for the G1 model (shown because of its higher resolution input data), a simple scatterplot of central pressure against maximum wind gust 'attributable' to the cyclone. The latter is represented, by proxy, as the maximum of the following: a 2-dimensional, gridded, near-circular, 600km-radius wind speed field, centred on the cyclone, in which every gridpoint value was computed, by vertical interpolation, to be 1km above the underlying model orography (or sea). In this way we are estimating gusts using mean speeds at the (approximate) top of the boundary layer.

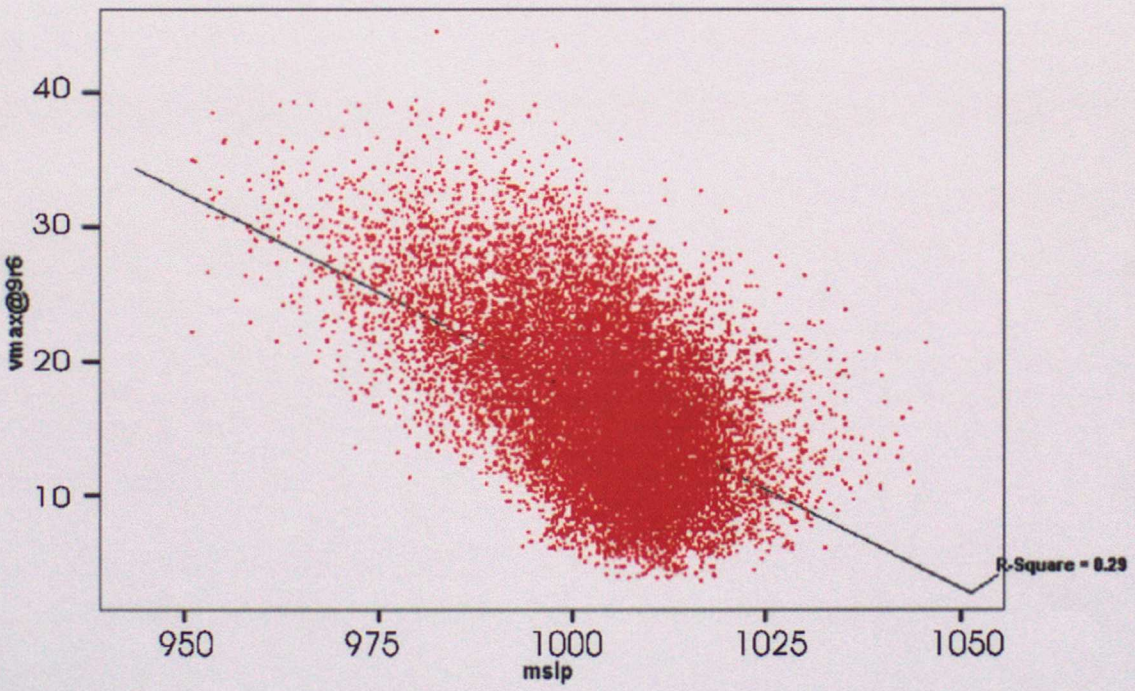

Figure 3: Scatterplot showing central pressure in mb (x) vs maximum estimated surface gust strength in m/s (y) likely to be encountered within a 600km radius of the cyclone centre, for the G1 runs during the trial. Estimated gust strength is given by mean wind speed interpolated to 1km above ground. A simple linear regression line is added.

There is indeed an overall increase in gust strength for deeper cyclones, but pressure only explains 29% of the gust variance. Significantly, a lot of the strongest gusts (around 40m/s or 80 knots) occur for central pressures between about 975 and 1000mb.

We now re-categorise cyclones using the gust-strength proxy variable, as was done in the G1 analysis for 2000 (see section 1). The top 2 panels of Figure 4 show maximum gusts for all cyclones. As with categorisation by central pressure there are again signs of under-prediction of the number of weaker cyclones in both models. Focussing on the 20-40 knot gust range (yellow and light grey) a shortfall of about 15% develops in both models by T+144. In G1 this predominantly occurs in the first 48 hours, as was the case in the 2000 analysis. However there

appears to have been some 'improvement' since 2000, when the quoted equivalent reduction was 25% (section 1). This may be a real and significant model improvement, but may alternatively be a sign of regime dependence of the error. Note that as a whole 2000 was unusually mobile, and cyclonic in the Northeast Atlantic, whilst the 2001 period studied, at least during November and December, was relatively blocked, with many lows tending to track north or northeast towards Greenland rather than east-northeast towards the UK.

Of passing interest also on the top left panel of Figure 4 is the marked 'diurnal' cycle seen in numbers of very weak cyclones (dark grey). This probably relates to heat lows (that form, for example, over Iberia). Had we composited model runs for both 00 and 12UTC data times this signal would probably not be visible (as is the case with the NT data).

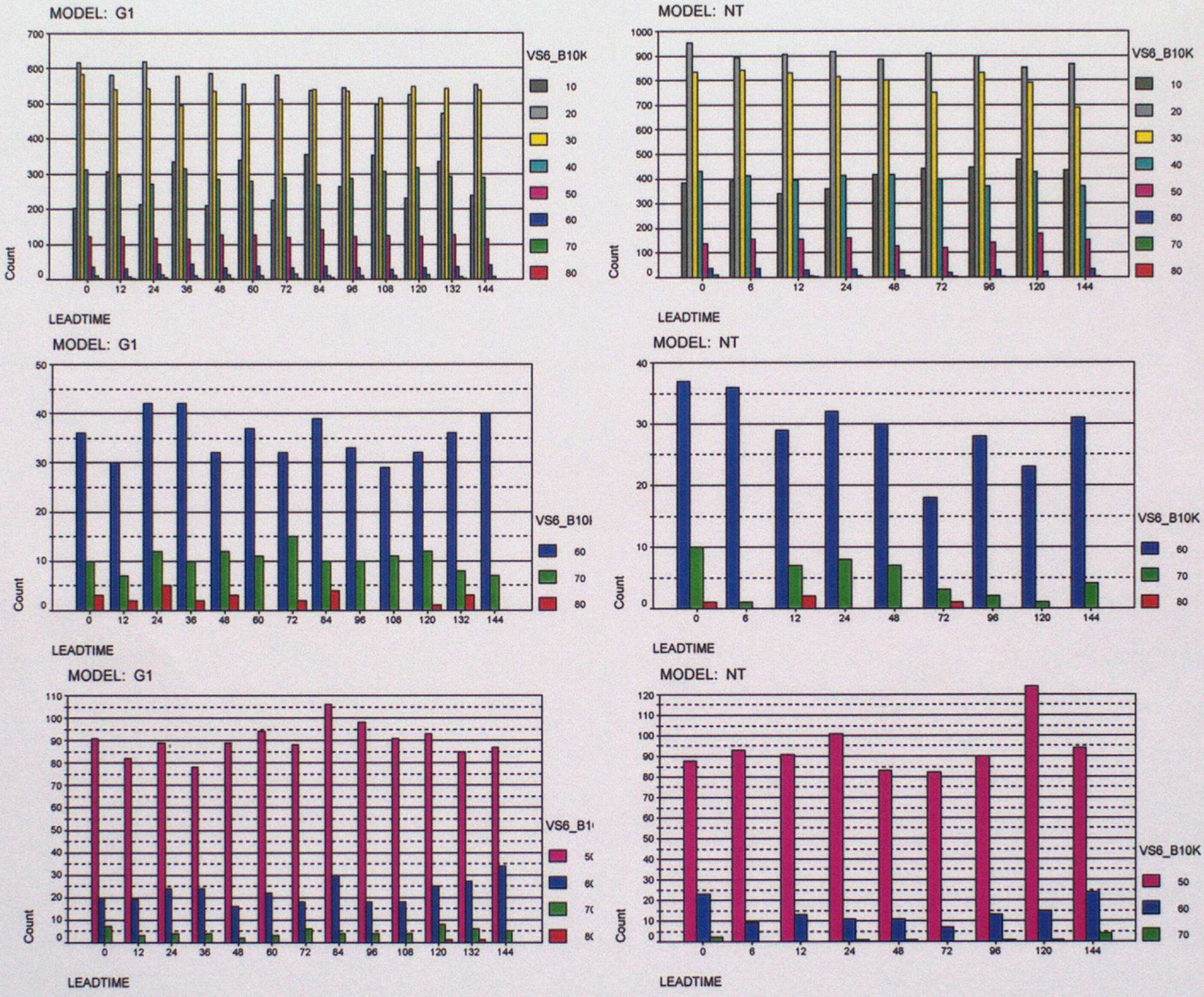

Figure 4: Cyclone spectra in G1 and NT runs as a function of lead time, with classification by maximum gust strength – using a proxy variable representing the maximum mean wind speed 1km above the earth's surface within 600km of the cyclone centre. First row shows all cyclones, second row only the more extreme cases. Third row is equivalent to the second except that cyclones in the Greenland area (shaded on Figure 5) are excluded. Legend numbers are the lower limits of 10kt ranges.

The second row of Figure 4 shows only the more extreme cyclones, with gusts above 60knots. Here the G1 model displays relatively stable behaviour, as it did in 2000, with no clear drift with lead time, and probably all the variations seen being attributable to sampling. This is optimal behaviour. Conversely there are signs in NT of a reduction with lead time in the numbers of extreme cases, which is of some concern. A particularly large drop, from 11 cases to 1, can be seen between T+0 and T+6 for cyclones producing gusts in excess of 70kts.

To further investigate the above we looked at case locations, and found that a lot of the strong wind producing cyclones were in the vicinity of Greenland. Figure 5 shows an example 'objective synoptic chart', taken from the forecast animation archive, on which the different cyclone types stored in the database are plotted as spots. Evidently, despite the high orography, the cyclone identifying methodology has made a good attempt at identifying sensible-looking cyclones in this general area. However this analysis includes also the most extreme NT cyclone case (gusts in excess of 80kts) – the feature, a frontal wave, is situated just southeast of Greenland (large orange spot), whilst the maximum gust attributed to it was over the higher ground of Greenland on the low's northwestern flank. This raises two issues. Firstly are we really interested in high wind strengths that are on occasion as much to do with the altitude of the underlying ground? And secondly do we want the proximity of such high ground to cause mis-attribution of a strong wind event to the wrong cyclone (on Figure 5 the strong winds are probably more closely related to another frontal wave west of Greenland)? To avoid these problems, which were compounded by the prevailing weather regime during the trial period, the dataset used in this study was pared down by removing all cyclones situated within the shaded area on Figure 5.

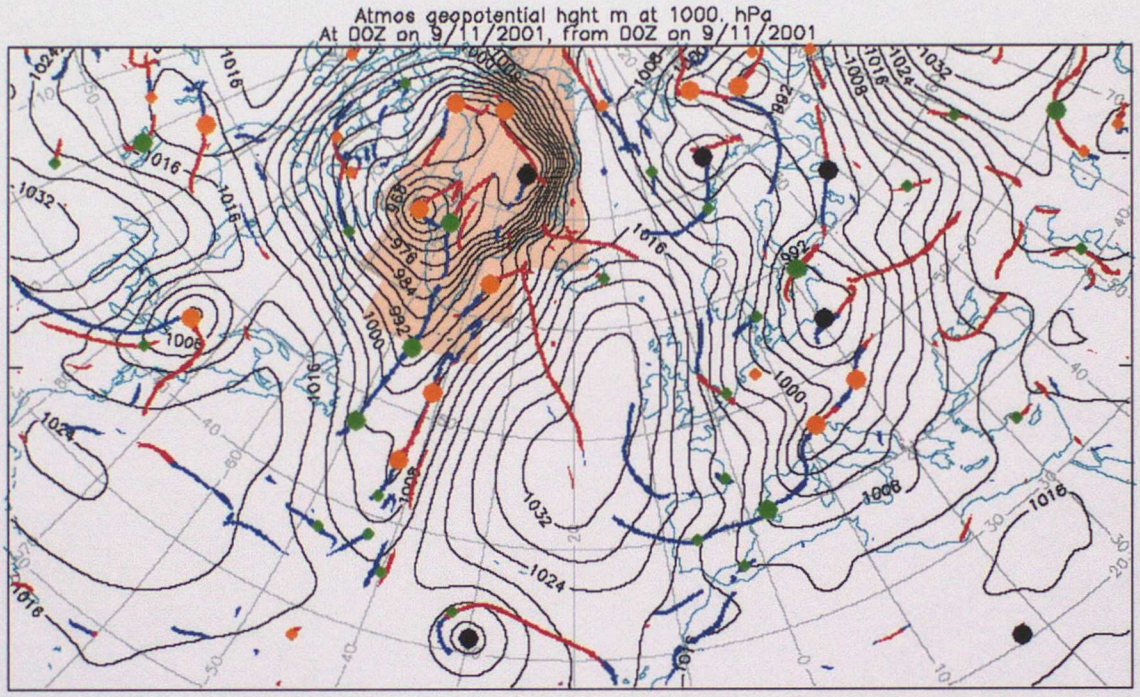

Figure 5: Objectively-analysed synoptic chart for 00UTC on 9 November 2001, for the NT model. Black contours are surface pressure as derived from 1000mb geopotential height. Red and blue lines are objective warm and cold fronts (with weak thermal threshold cut-offs applied). Standard frontal waves are shown by large orange spots and barotropic lows by black spots – these are the 'significant cyclones' analysed in this report. Large green spots are potential waves. Smaller green and orange spots are weaker waves. The shaded area is the 'Greenland area', excluded from some of the analyses in this report.

On Figure 4 the third row is equivalent to the second row but with the Greenland cases (approximately 50%) removed, and with a 50 to 60kt gust category added. In broad terms the overall signal is unchanged, except that now there are signs, for both models, of a recovery/increase in numbers of extreme cyclones around days 5 and 6. Between T+6 and T+96 in the NT runs there are still 50% fewer cyclones associated (nominally) with gusts greater than 60 knots than in the NT analyses. The sharp drop off between T+0 and T+6 has been investigated further.

Figure 6 shows the locations of all intense NT cyclones, at T+0 (solid) and T+6 (open), as represented on the lower right panel of Figure 4 (same colours). We will consider just the cases greater than 60 knots (blue and green). In the northeast of the chart there are clearly far more cases analysed (7) than found in 6-hour forecasts (1); this looks significant. In the western North Atlantic 4 cases (labelled) randomly selected from those plotted on Figure 6 were investigated more thoroughly, using the archived animations. In 2 the cyclone weakened between T+0 and

T+6, in the others it intensified. However when each T+12 frame was then validated against the next NT analysis, in every case the cyclone was too weak. Together with Figure 4 these observations suggest there may be a spin-up problem in the NT model.

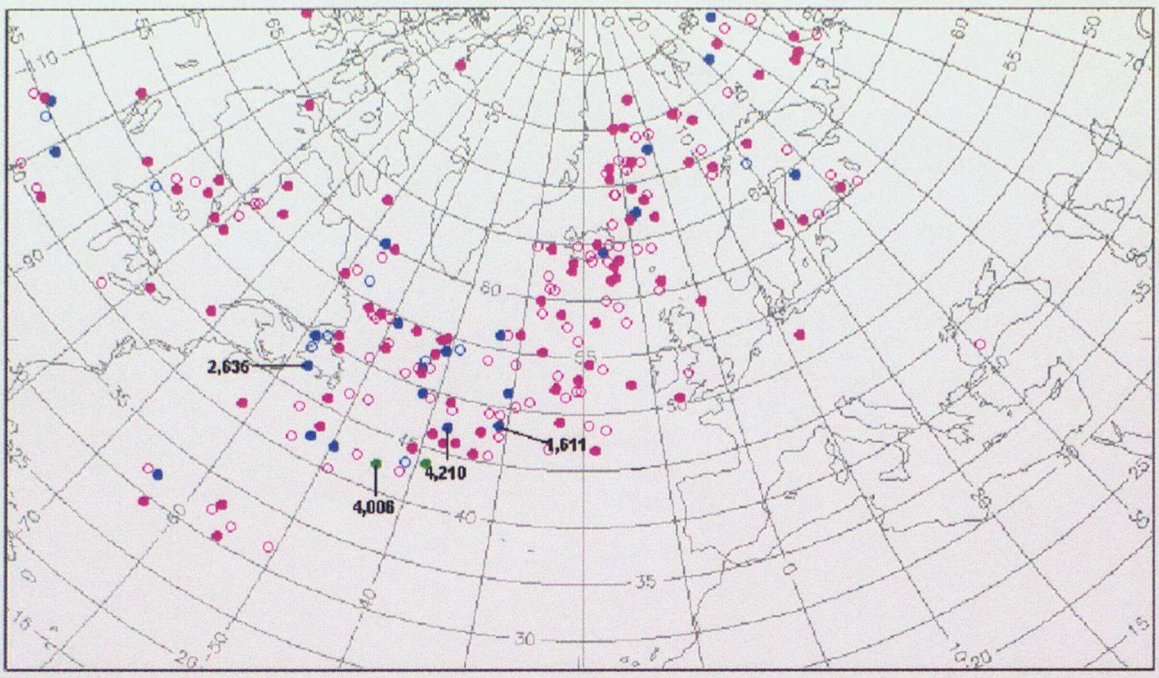

Figure 6: NT Cyclones at T+0 (solid) and T+6 (open) which were nominally associated with maximum gusts in the range 50-60 kts (pink), 60-70kts (blue) and 70-80 kts (green). Cyclones in the Greenland area (shaded on Figure 5) have been excluded. Labelled cases were investigated in more detail (see text) – case 1611 is shown on Figure 7.

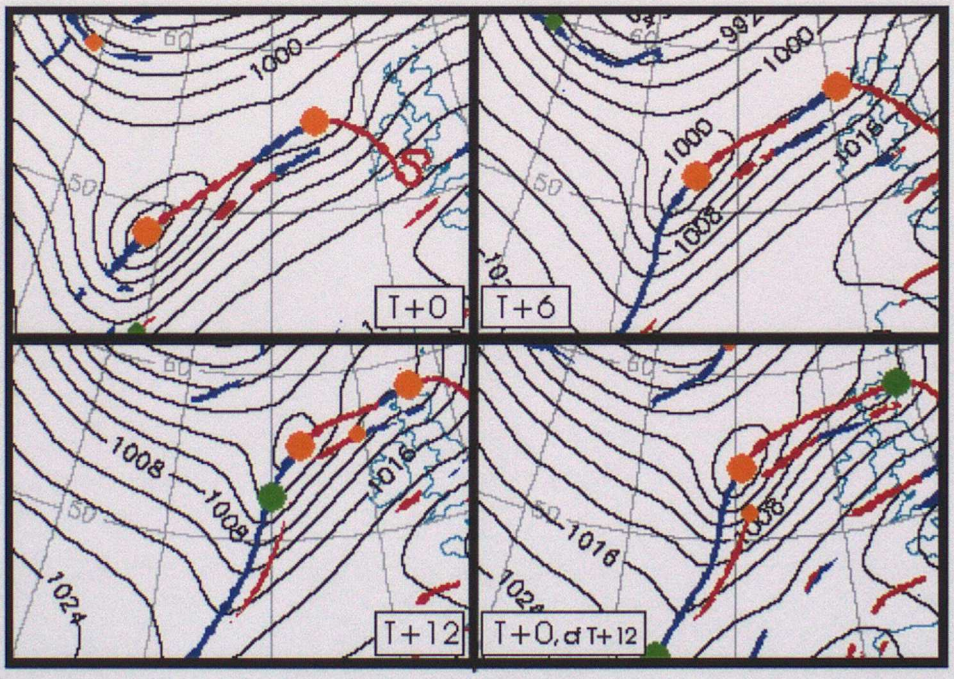

Figure 7: An example of the NT model weakening a cyclone that should have been retained as an intense feature, showing surface pressure, objective warm and cold fronts, frontal wave cyclones (orange spots) and potential waves (green spots). The first T+0 frame is for 00UTC on 10th October 2001.

To illustrate the type of synoptic scale model evolution we may be seeing, Figure 7 shows, for a mid-range case of the four analysed above, the T+0, T+6 and T+12 frames, together with the analysis that validates at T+12. The frontal wave near the southwest of the first chart is asymmetrical, having strong gradients on its southern and southwestern flanks. By T+12 this has moved northeast but weakened somewhat, into a more elongated low (as denoted in part

by the appearance of a potential wave at this time). However the validating analysis for this time (lower right panel) depicts a low on which the strong gradients on the southern flank have been retained. It is in exactly this part of the system that damaging gusts we see from time to time over the UK tend to occur. Similarly it is critically important (though nonetheless difficult!) to predict whether such a wave will weaken, as this rather poor forecast suggested it would, or retain its strength or intensify. Note also that the wave's central pressure differs little on the lower two panels, showing again how Figures 2 and 4 can give different though not necessarily inconsistent impressions of a model's ability to correctly handle cyclone development.

Without further investigation it is impossible to say how often or how serious an error that shown by Figure 7 might be, and indeed there are plenty of other ways, synoptically, that a low can be made too weak in the first few hours of a forecast. Similarly there will doubtless also be other occasions when a low is forecast to intensify by the NT model, when in reality it doesn't. However the bulk statistics presented so far do seem to suggest some sort of NT model problem arising early in the forecast period, which will manifest itself on occasions as an error such as shown by Figure 7. Conversely, because of a lack of T+6 G1 data, it is not possible to say whether the G1 model has a similar spin up problem, although given that the differences seen between T+0 and T+12 in NT are not apparent for G1 (lower panels of Figure 4) this seems unlikely.

Deepening of frontal wave cyclones often depends on the interaction between upper and lower level anomalies. When this interaction is occurring there will be considerable ascent within the cyclonic system, and similarly compensating descent just upstream of its centre. Indeed it is often the closeness of these regions that can help strengthen pressure gradients and thus give rise to very strong winds at the surface. The forecast example in Figure 7 seems to be lacking or dampening this interaction, for whatever reason.

A good measure of the forcing of vertical motion in the atmosphere can be obtained by inverting a dry dynamics version of the omega equation (Clough et al (1996), Clough and Davitt(1994)) to develop so-called attribution diagnostics (giving ω_{qg}, or a 'quasi-geostrophic omega'). One can also partition ω_{qg} into components that are due to forcing by upper and lower levels. Because of its integrated properties, ω_{qg} seems to be a good bulk candidate for evaluating the tendency of a model to produce cyclonic systems, and for assessing any drifts in such a tendency.

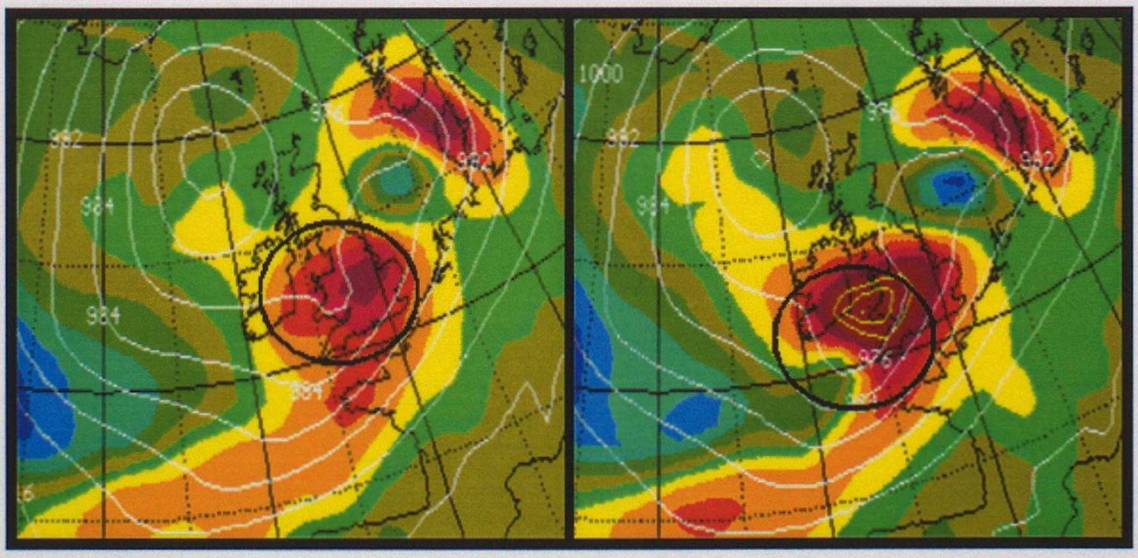

Figure 8: Surface pressure (contours) and ω_{qg} at 700mb (coloured) for 0300UTC on 30th October 2000. Colour bands cover 1cm/s intervals, with zero at the green/yellow boundary. The left panel is from a poor, insufficiently developmental forecast, whilst the right panel is from a much better forecast. Rings with a 300km radius are centred on the developing storm system.

Figure 8 shows ω_{qg} at 700mb (coloured) and pressure (contours) for the major storm of 30th October 2000 (ringed) about 12 hours before it reached maximum intensity (Hewson (2001b)). The left panel is taken from a poor forecast, which suffered from similar (though more substantial) problems to those highlighted by Figure 7, whilst the right panel is from a good

forecast. There are two key differences in the ω_{qg} structures, which relate to the subsequent forecast evolution. On the good forecast (right) the maximum value, within the ring, is higher (+9cm/s compared to +5cm/s on the poor forecast), and there is a clear minimum value on the southwestern flank of the low (-3cm/s compared to +1cm/s on the poor forecast). Deveson et al (2002) also discuss these forcing couplets, and how they can be used to classify cyclones.

A wide range of ω_{qg} diagnostics is stored in the cyclone database. We now use these to look at the models' tendency to produce and intensify cyclones. Figure 9 shows the distribution, by percentage population, of the maximum (top half) and minimum (bottom half) ω_{qg} values found for all cyclones within 300km radii of their centres (ie within rings as on Figure 8). Cyclones that show a greater propensity for intense development should populate the extreme ends of the figure – most notably the top, but if strong winds are to develop then also, by analogy with Figure 8, the bottom. On the NT plot the smooth and consistent expansion of the central bands, at the expense of the outer bands seems to clearly show a trend, with forecast lead time, to produce cyclones that have a lower propensity to develop strongly. This is even allowing for the fact that we have fewer cyclones anyway at the longer lead times (Figures 2 and 4). The largest jump occurs between T+0 and T+6, which seems to concur with earlier comments about a probable, if rather complex, spin up problem. For the G1 model there also appears to be a slight though lesser trend in the same direction (about 4 and 5% increases respectively, in the size of the green and grey bands from T+0 to T+120, compared to about 6 and 8% for NT), albeit with some apparent 'return of forcing' around days 5 or 6 – consistent perhaps with trends hinted at at this range on Figures 2 and 4. There are also signs again of a diurnal cycle in the minimum values. Whether this relates to heat lows or could be a problem in itself is not yet clear. For fair comparison we should also divide up the NT forecasts, into 00z and 12z data times, to allow any equivalent signal to show through. Note also that the differences in vertical resolution of the input data for G1 and NT (Section 2) should have had only a small impact on ω_{qg}.

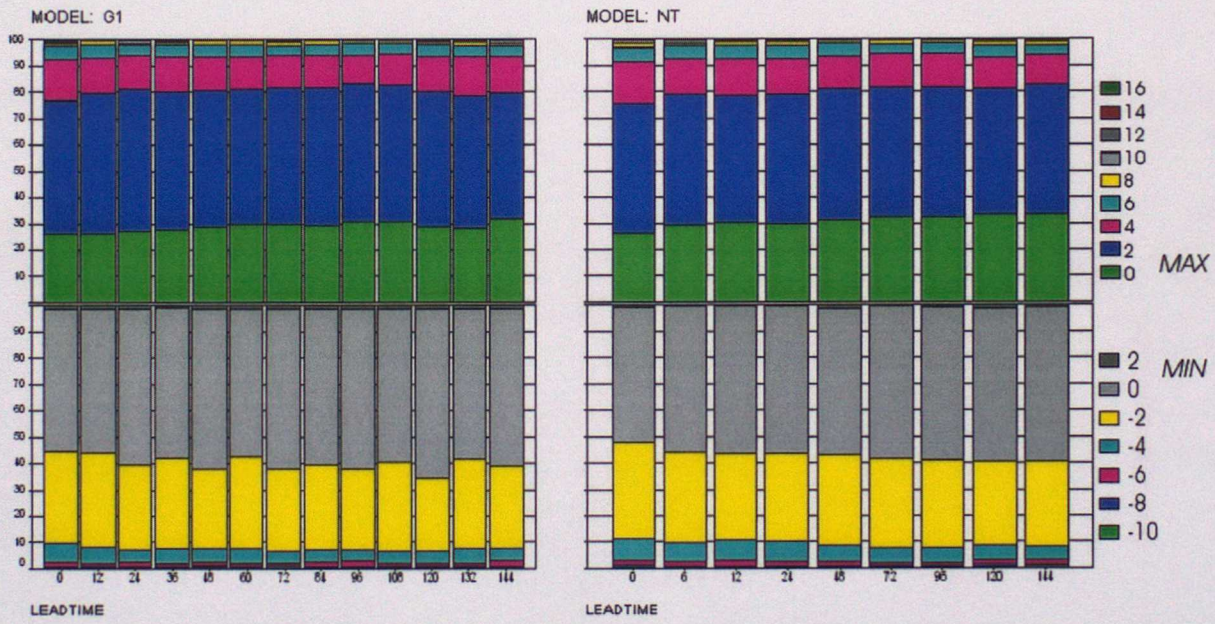

Figure 9: Top half shows distribution by percentage of the maximum value in cm/s of total 700mb ω_{qg} found within a 300km radius of each cyclone centre, as a function of forecast lead time, for all cyclones excluding those in the Greenland region (shaded on Figure 5). Bottom half is similar, but represents the minimum ω_{qg}.

To examine whether the trends seen in Figure 9 perhaps have their origins at upper levels we show in Figure 10 those cyclones that are associated with a significant upper level forcing dipole – ie 'dipole strength' greater than 1cm/s. Upper level dipole strength is calculated using the component of ω_{qg} at 700mb that is attributable to thermodynamic forcing from levels in the atmosphere above 650mb; it represents the mean magnitude of the maximum and minimum values of this variable found within 600km radii of a cyclone centre (600km being more appropriate than 300km for capturing the larger scale upper-level extrema). Overall, on Figure 10, there seems to be a marked reduction, of say 40% by T+96, in the number of cyclones in this category in the NT runs, although there isn't such a sudden change as seen for other measures between T+0 and T+6, suggesting that the apparent low level spin up problem may

be more to do with low level *response* to upper level forcing as opposed to a problem triggered by deficiencies at the upper levels themselves. The plot for G1 shows a very different and rather surprising structure whereby upper level forcing around cyclones drops markedly in the first 12 hours, but then proceeds to recover gradually through the rest of the forecast, perhaps increasing too much towards the end given the disproportionate numbers of more intense cyclones at days 5 and 6 that were seen in various other measures.

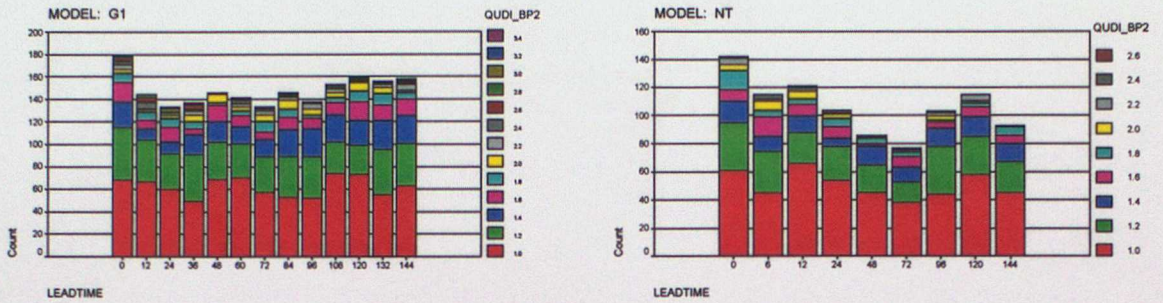

Figure 10: Frequencies of cyclones associated with an upper level forcing dipole strength of greater than 1cm/s (see text for explanation). Colours show this dipole strength. Cyclones in the Greenland area (shaded on Figure 5) are excluded.

One aspect that analysis of QG forcing tends to miss is the impact of latent heating on cyclone strength, because dry dynamics are assumed. Deficiencies in latent heating could in principal be enough to greatly amplify, or indeed offset, the signals described above, although the fact that the trends we are seeing tend to be mirroring those seen in the wind and pressure diagnostics tends to suggest that the latent heating is being handled adequately in both models.

A limited investigation of the geographical distribution of the 15% shortfall in numbers of weaker cyclones noted on Figure 4 has also been performed, looking at the area 90W to 50W, 40 to 45N which on the 2000 G1 analysis was a significant region of under-prediction (Figure 1). In neither the NT nor G1 models was there was any sign of a marked reduction in numbers here with lead time. This contrasts strongly with the 2000 results, although it is not clear whether it signifies a modelling improvement or a regime dependant error.

Given that changes were being continually introduced into the NT model formulation throughout the trial period, it is possible that some of the weaknesses and trends in the handling of cyclones that have been described in this report no longer exist. The trial period was divided up into two epochs (yellow line on table 1) to see if any trends due to model changes could be identified. However this part of the study gave no conclusive results, partly because of the shortness of the two periods, but mainly because the second, being more a winter season, contained the vast majority of the more intense cyclones. However if nothing else this latter observation does suggest that most of the signals noted concerning intense cyclones were probably present late in the period, and so probably still need addressing. Tests were also performed to see if any of the results, which were intended to relate to mid-latitude cyclones, had been 'contaminated' by tropical or sub-tropical systems. Because of the noted improved handling of tropical systems by the NT model, and the relatively low latitudes to which the southwestern part of the LAM domain extends (see Figure 1) this was seen as important. Splitting the database with a cyclone-centred 1000-500mb thickness cut-off of 558dm, to isolate what might potentially be systems of tropical origin, showed up no major departures from the trends noted in this section.

4. Conclusions

The main results of this study can be summarised as follows (NT = New Dynamics, G1 = current operational UM):

• The tendency noted in G1 in 2000 to under-predict the total number of weak cyclones (mostly frontal waves) occurred again in the trial in both G1 and NT. Both showed about a 15% reduction in numbers through the forecast, which seems better than the 25% reduction noted in

2000. However the shortfall did not occur in the same locations as in 2000, perhaps because different weather regimes prevailed.

• The G1 model continues to show a consistent handling of more intense cyclones through the forecast period, with no major biases apparent, as was the case in 2000, although towards days 5 and 6 there may now be a slight tendency to over-intensify some systems.

• The NT model seems to be *slightly* inferior to the G1 model in its handling of intense systems - there is something of a downward trend in numbers with forecast lead time. This does not show up in cyclone centre mslp, but is apparent in a more relevant diagnostic showing maximum wind speeds attributable to a cyclone. In turn this suggests that the pressure pattern around NT cyclones is, on average, insufficiently asymmetric. Of note here is the fact that the semi-Lagrangian scheme used to time-integrate the NT model has some intrinsic smoothing associated with it, which probably tends to make systems slightly too symmetric.

• Much of the downward trend in numbers of extreme cyclones in NT occurs in the first 6 hours of the forecast, suggesting some sort of spin up problem. This may perhaps be an artifact of the analysis system that is currently tuned for the G1 model. However it is potentially a serious problem for shorter range forecasting in the vicinity of the UK, where the mesoscale model may not be able to overcome the detrimental impact of forcing by inaccurate boundary conditions.

• Diagnostics representing quasi-geostrophic forcing around the analysed cyclones showed a small but clear and consistent trend, throughout the NT runs, towards less developmental patterns. G1 showed a similar but smaller and more erratic signal. This suggests that diffusion introduced into the NT runs, primarily to quell convective instabilities, could perhaps be reduced somewhat until a zero trend is seen, provided this had no other adverse impact, although it seems probable that there are some other causes. Work is in progress to try to reduce the need for the diffusion.

These results should be accompanied by a number of 'health warnings'. Firstly, one must always be a little cautious about treating T+0 fields as truth, because of their dependance on the assimilation scheme. Secondly, continual model updating, whilst undoubtedly a practical necessity, can of course adversely impact on any analysis performed over a long period, and it is of course necessary to use a reasonably long period to capture sufficient of the more extreme events. Indeed the ideal would be to use several years' data, including other seasons, and including other regimes. Evidently any real- time test is at the mercy of the prevailing weather type. Despite having looked, in this study, at about 50,000 instances of cyclones appearing in model fields, there is still some uncertainty in the results; there is a clear message here regarding the relevance of individual case studies.

Overall, in the author's view, the deficiencies noted in the NT runs are not that severe, and are probably counterbalanced by the better rms scores that it achieves relative to G1. However the apparent spin up problem does seem to warrant close attention. The facility for performing stringent testing, of a similar type, should be retained, especially when any significant model changes are introduced, such as a reduction in the diffusion. If at all practicable, tests should carried out over a longer period than has been used here. Greater use of tests of statistical significance is also to be encouraged.

Acknowledgements:

Thanks to Andy White, Sean Milton and Nigel Wood for helpful comments on an earlier draft of this report.

References:

Clough, S.A. and Davitt, C.S.A. 1994. 'Development of an Atlantic frontal wave during IOP3 of Fronts92. Pp 151-156 in *The life cycles of extra-tropical cyclones; vol II*. Eds S. Gronas and M.A.Shapiro, Geophysical Institute, University of Bergen.

Clough, S.A, Davitt, C.S.A and Thorpe, A.J., 1996. Attribution concepts applied to the omega equation. *Q.J.R. Meteorol. Soc.* **122**, 1943-1962.

Deveson, A.C.L., Browning, K.A., Hewson, T.D. 2002 'A Classification of FASTEX cyclones using a height-attributable quasi-geostrophic vertical-motion diagnostic'. *Q. J. R. Meteorol. Soc.* **128**. 93-117.

Hewson, T.D. 1997. 'Objective identification of frontal wave cyclones' *Meteorol. Appl.* **4**. 311-315.

Hewson, T.D. 1998a. 'Objective fronts'. *Meteorol. Appl.* **5**. 37-65.

Hewson, T.D., 1998b. 'A Frontal Wave Database'. JCMM Internal Report No. 85. pp 20. Department of meteorology, Reading university.

Hewson, T.D., 2001a. 'A Cyclone Database'. Met Office Forecasting Products division, Project Report. 20 pages.

Hewson, T.D., 2001b. 'The North Sea Storm of 30th October 2000', Weather. **56**. 115-116.